



Statistiques et motifs séquentiels : le cas des suites [Nom préposition Nom], mot à mot, pétale par pétale

Véronique Magri-Mourgues

► To cite this version:

Véronique Magri-Mourgues. Statistiques et motifs séquentiels : le cas des suites [Nom préposition Nom], mot à mot, pétale par pétale. XVIIe Congrès de linguistique et de philologie romanes, Jul 2016, Nancy, France. pp.175-188. hal-01372923

HAL Id: hal-01372923

<https://hal.science/hal-01372923>

Submitted on 27 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistiques et motifs séquentiels : le cas des suites [Nom préposition Nom], *mot à mot, pétale par pétale*

Véronique Magri-Mourgues

Univ. Nice Sophia Antipolis, CNRS, BCL, UMR 7320, 06300 Nice, France.

Résumé

La communication se propose d'évaluer l'outil statistique pour identifier et analyser les suites *Nom préposition Nom*, avec pour critère restrictif que les deux substantifs sont identiques. L'enjeu essentiel de cette recherche est de montrer que ce type de structures peut être utilisé à des fins de classifications.

La base qui sert de corpus exploratoire à cette recherche est constituée par un extrait de *Frantext* catégorisé : il comprend quatre groupes génériques répertoriés dans *Frantext* : d'une part deux ensembles narratifs, le roman et les mémoires, d'autre part la poésie et le théâtre afin que les genres principaux soient tous représentés.

Le repérage des unités *NpN* permet une réflexion sur le statut de la préposition, à la fois sur le plan lexical et sémantique, et permet de réfléchir au partage entre structures figées et innovations lexicales, entre langue et discours. Ces unités peuvent être proposées comme indices d'un style d'auteur ou même comme marqueur générique.

1. Introduction

La communication se propose d'évaluer l'outil statistique pour identifier et analyser les suites *Nom préposition Nom*, avec pour critère restrictif que les deux substantifs sont identiques. L'enjeu essentiel de cette recherche est de montrer que ce type de structures peut être utilisé à des fins de classifications.

La base qui sert de corpus exploratoire à cette recherche est constituée par un extrait de *Frantext* catégorisé ; il comprend quatre groupes génériques répertoriés dans *Frantext* : deux ensembles narratifs, les mémoires (109 textes) et les romans (740 textes) d'une part, la poésie (191 textes) et le théâtre (210 textes) d'autre part, tels qu'identifiés dans la base, soit près de 85 millions de mots. Les genres principaux sont ainsi représentés.

Le premier travail consiste à repérer les suites recherchées. Il s'avère que toutes les prépositions ne sont pas aptes à entrer dans ce type de construction. Seules seront retenues *à, après, contre, par, pour, sur* ainsi que les structures complexes *de...à* et *de...en*. Une réflexion pourra s'engager sur le statut des prépositions et sur leur portée sémantique. Le rendement des structures est évalué en observant les variations de substantifs de part et d'autre des prépositions : la préposition peut varier tandis que les noms restent identiques et l'expression conserve globalement le même sens (*mot à mot, mot pour mot ; bouche à bouche, bouche contre bouche* ; la préposition peut être la même tandis que les substantifs peuvent varier (*nuit après nuit, soir après soir* : le sens exprime ici l'idée de succession ;

rendre haine contre haine, corps à corps, nez contre nez : le sens se module en l'idée d'échange, d'opposition, de contact).

L'analyse a aussi à composer avec les structures lexicalisées ; on s'intéressera ainsi aux critères et aux fonctions statistiques qui autorisent une interprétation sur le degré de figement des structures. Entre phraséologie et style d'auteur, conçu comme appropriation des unités linguistiques, entre langue et discours, ces unités seront évaluées en fonction de leur productivité dans le corpus d'étude. L'observation des hapax – ou unités de fréquence 1 – sera engagée comme possible critère distinctif de l'innovation lexicale. L'étude amorcera par ce biais une délimitation et une possible définition des unités lexicalisées.

On envisagera par ailleurs l'environnement cotextuel des unités considérées afin d'observer la fonction syntaxique assurée dans la phrase ainsi que la place occupée dans la séquence textuelle. On verra si une fonction structurante peut leur être attribuée.

Comme le corpus est constitué de quatre ensembles génériques, il sera loisible de voir enfin comment ces unités phrastiques se distribuent et si une répartition générique peut être proposée.

- I. L'unité *Nom préposition Nom*
- II. Fréquence et productivité des unités
- III. Les unités et les genres textuels

I. L'unité *Nom préposition Nom*

- Observation du corpus

On commence par l'observation des données expérimentales fournies par le corpus d'étude afin de dégager quelles sont les prépositions rentables, autrement dit susceptibles de relier immédiatement deux substantifs identiques, chacun étant dépourvu de déterminant.

Dans le corpus d'étude, des occurrences de la suite ternaire recherchée ne se rencontrent qu'avec les prépositions : *à, après, contre, par, pour, sur* ainsi que les structures plus complexes *de ... à* et *de ... en*¹.

Effectifs :

Structure = N préposition N	Total	Exemple d'expression	Structure = N préposition N	Total	Exemple d'expression
à	2997	<i>mot à mot</i>	d(e) ... en	4222	<i>de page en page</i>
après	247	<i>heure après heure</i>	par	969	<i>goutte par goutte</i>
contre	253	<i>cœur contre cœur</i>	pour	398	<i>coup pour coup</i>
d(e) ... à	372	<i>d'ami à ami</i>	sur	663	<i>pierre sur pierre</i>
			Total	10121	

¹ La préposition *selon* ne fournit qu'une seule occurrence : le « pli selon pli » de Mallarmé dans le sonnet « Remémoration d'amis belges » : « Que se devêt pli selon pli la pierre veuve ».

Est appelée *expression ou unité* toute réalisation d'une structure NpN à partir d'un substantif particulier : *heure après heure, d'ami à ami*. Sur le plan terminologique, on pourrait réutiliser le terme de "motif", défini de manière un peu abstraite dans certains travaux comme « un sous-ensemble ordonné [d'un ensemble] (E) formé par l'association récurrente de n éléments de l'ensemble (E) muni de sa structure linéaire » (Longrée, Luong, Mellet 2008).

Sur le plan formel, les structures NpN peuvent entrer dans la catégorie du motif syntaxique, qui est une des catégories admises du motif, à côté de celles constituées par les suites phonologiques ou encore métriques. La suite choisie présente la particularité que les noms sont en théorie réversibles puisqu'ils sont identiques dans leur forme ; cependant, l'ordre d'apparition des occurrences dans la phrase a forcément une incidence sur le sémantisme du lexème. Son inscription dans la linéarité de la phrase induit une variation de sens d'une occurrence à sa répétition. L'agencement des unités, leur disposition dans l'espace du texte donne une orientation à la structure en l'inscrivant dans la temporalité, qui est d'abord celle de la lecture cursive de la structure.

La structure choisie se situe entre langue et discours. Le chercheur doit composer avec les lexies ou unités phraséologiques introduites dans le discours et les créations de séquences ternaires sur un patron linguistique préconstruit. L'analyse croise des structures qu'on pourrait dire figées puisqu'elles sont constituées en langue, comme syntagmes adverbiaux. La lexicalisation de la structure se vérifie notamment lorsqu'elle repose sur un transfert normalisé du sens propre au sens figuré :

Ce n'est pas que Combeferre ne fût capable de combattre, il ne refusait pas de prendre *corps à corps* l'obstacle et de l'attaquer de vive force et par explosion. (V. Hugo, *Les Misérables*)

Dans d'autres exemples, le jeu sur l'emploi du sens concret ou figuré de la structure lui redonne son emploi discursif susceptible de créer un effet.

[1] Tout le monde descend l'escalier ; toi, tu déchireras tes draps de lit, tu en feras, *brin à brin*, une corde, puis tu passeras par ta fenêtre, et tu te suspendras à ce fil sur un abîme. (V. Hugo, *Les Misérables*)

[2] Aux quatre vents, quelle tempête s'amasse sur le lac du genre humain ! N'est-ce pas la création sans foi qui se détache *brin à brin* des mains du créateur, et tombe dans l'abîme, comme le chapelet d'un prêtre d'Arménie tombe à ses pieds. (É. Quinet, *Ahasverus*)

La contextualisation de la structure permet d'appréhender celle-ci : elle est susceptible d'être envisagée comme lexie quand l'emploi métaphorique est neutralisé par l'usage comme dans l'exemple [2]. Sa place entre langue et discours est cependant préservée dans tous les cas car le contexte est toujours à même de remotiver une figure lexicalisée.

Les critères définis dans la définition donnée du motif associent à des paramètres quantitatifs comme la fréquence des propriétés textuelles fondées sur la récurrence qui doit permettre la mémorisation du motif et son rôle dans l'agencement discursif susceptible de caractériser certains usages.

Le corpus d'étude fournit 10 122 occurrences de la structure ternaire choisie. Indépendamment de la partition générique, les expressions les plus fréquentes sont :

côte à côte	1107	nez à nez	165
goutte à goutte	359	jour par jour	150
coup sur coup	315	de main en main	133

de minute en minute	223	de porte en porte	124
corps à corps	192	tête à tête	108
de place en place	186		

- Les unités et le paramètre sémantique

Les prépositions sont, pour la plupart, polyvalentes et polysémiques. C'est le sens des substantifs environnants qui détermine le sens de la relation sémantique instaurée entre les noms et, par conséquent, le sens de l'expression.

L'examen détaillé des expressions NpN du corpus permet d'établir une liste de huit structures différenciées par le paramètre sémantique² :

Succession : décomposition du procès en éléments

Nous y secouons nos cigarettes: nous sommes lavés, nous avons le temps ; le temps chaud et stagnant , le temps qui passe *minute à minute*, sans bruit, sans fièvre, en chuchotant. (A. Sarrazin, *L'Astragale*)

Accumulation : procès d'addition d'objets

La mère, toujours malade, comptait *sou à sou*. Elle tremblait devant lui, autant que nous. (R. Martin du Gard, Les Thibault, l'Été 1914)

Mon grand vizir est à la guerre où il remporte d'ailleurs *victoire après victoire* sur les Russes (M. De Grèce, *La Nuit du Sérail*)

Distance : topographique ou chronologique

J'ai tendu des cordes *de clocher à clocher* ; des guirlandes *de fenêtre à fenêtre* ; des chaînes d'or, *d'étoile à étoile*, et je danse. (A. Rimbaud, *Illuminations*)

Déplacement - mouvement

Je marche *de surprise en surprise*, *d'ahurissement en ahurissement*. (P. Benoit, *L'Atlantide*)

Relation : entre sujets (dont synecdoque)

D'où Nana tombait-elle ? Et des histoires couraient, des plaisanteries chuchotées *d'oreille à oreille*. (É. Zola, *Nana*)

Mille cris se croisaient *de baraque à baraque*. (P. Adam, *L'Enfant d'Austerlitz*)

Contact : entre parties de sujets

J'ai l'horreur de la fraternité qui établit des contacts *de peau à peau* (J. Genet, *Miracle de la rose*)

Disait Hérode à Sigognac, qui cheminait *botte à botte* avec lui (Th. Gautier, *Le Capitaine Fracasse*)

Opposition : entre parties de sujets

On se battit *corps à corps* sous les arbres (V. Hugo, *Les Misérables*)

J'ai lutté pendant quatre années jour et nuit, *corps à corps*, *astuce contre astuce*, avec ce génie infernal (Ponson du Terrail, *Rocambole*)

Échange : échange d'objets

Violence contre violence. Chacun donne ici sans mesure sa force. (Aragon, *Les beaux Quartiers*)

² Sur 10121 occurrences, 11 n'entrent pas dans ces 8 catégories et sont donc soustraites des effectifs.

Structure	à	après	Contre	de ... à	de ... en	par	pour	sur	Total
Accumulation	5	28	2	0	0	15	0	253	303
Contact	1790	0	190	2	0	0	0	51	2033
Déplacement	0	0	0	25	4222	0	0	0	4247
Distance	1	0	0	6	0	0	74	0	81
Échange	0	0	11	0	0	6	324	0	341
Opposition	171	0	50	0	0	0	0	0	221
Relation	30	0	0	329	0	0	0	0	359
Succession	994	219	0	5	0	948	0	359	2525
Total	2991	247	253	367	4222	969	398	663	10110

Tableau des fréquences selon les structures et les sémantismes

On observe la polysémie des structures ou leur équivalence sémantique.

En conservant soit la préposition, soit le substantif comme invariant d'une structure, et en faisant varier l'autre élément, on obtient les deux cas de figure suivants qui illustrent un processus interactif de constitution du sens :

- ✓ Une structure qui combine une même préposition avec des noms différents peut se charger de sens différents, preuve que la préposition n'a pas un sens absolu mais toujours dépendant des combinaisons où elle entre. Son sens se module en interaction avec le substantif répété de part et d'autre. La préposition « contre » par exemple peut exprimer trois types de relations sémantiques : l'*Opposition* : « acier contre acier » ; le *Contact* : « bouche contre bouche » ; l'*Échange* : « rendre haine contre haine ».

La structure N *contre* N peut être dite polysémique.

- ✓ De manière symétrique, des substantifs identiques reliés par des prépositions différentes peuvent constituer une structure de même sens : par exemple, entre les expressions *lutter âme à âme* et *lutter âme contre âme* ou encore entre *bouche à bouche* et *bouche contre bouche* peut être postulée une équivalence de sens. Un même sémantisme est réalisé par plusieurs structures. Les statistiques sont maintenant sollicitées pour valider ou non ces hypothèses de travail intuitives.

Structure	à	après	contre	d(e) ... à	d(e) ... en	par	pour	sur	Total
Accumulation	0,17%	11,34%	0,79%			1,55%		38,16%	3,00%
Contact	59,85%		75,10%	0,54%				7,69%	20,11%
Déplacement				6,81%	100,00%				42,01%
Distance	0,03%			1,63%			18,59%		0,80%
Échange			4,35%			0,62%	81,41%		3,37%
Opposition	5,72%		19,76%						2,19%
Relation	1,00%			89,65%					3,55%
Succession	33,23%	88,66%		1,36%		97,83%		54,15%	24,98%

Tableau des pourcentages des sémantismes selon les structures

Ce tableau présente les différents sémantismes associés à une même structure.

– Les structures les plus polysémiques sont *à* (6 sémantismes lui sont associés), *de ... à* (5) et *contre* (4).

- En revanche, *De ... en* est monosémique et indique essentiellement le déplacement.
- Chaque structure a un sémantisme nettement dominant, intrinsèque, sauf *à* (*Contact* et *Succession*), *contre* (*Contact* et *Opposition*) et *sur* (*Accumulation* et *Succession*).

Structure	à	après	contre	d(e) ... à	d(e) ... en	par	pour	sur
Accumulation	1,65%	9,24%	0,66%			4,95%		83,50%
Contact	88,05%		9,35%	0,10%				2,51%
Déplacement				0,59%	99,41%			
Distance	1,23%			7,41%			91,36%	
Échange			3,23%			1,76%	95,01%	
Opposition	77,38%		22,62%					
Relation	8,36%			91,64%				
Succession	39,37%	8,67%		0,20%		37,54%		14,22%

Tableau des pourcentages des structures selon les sémantismes

Ce tableau présente les différentes réalisations concrètes (les structures) associées à un même sémantisme.

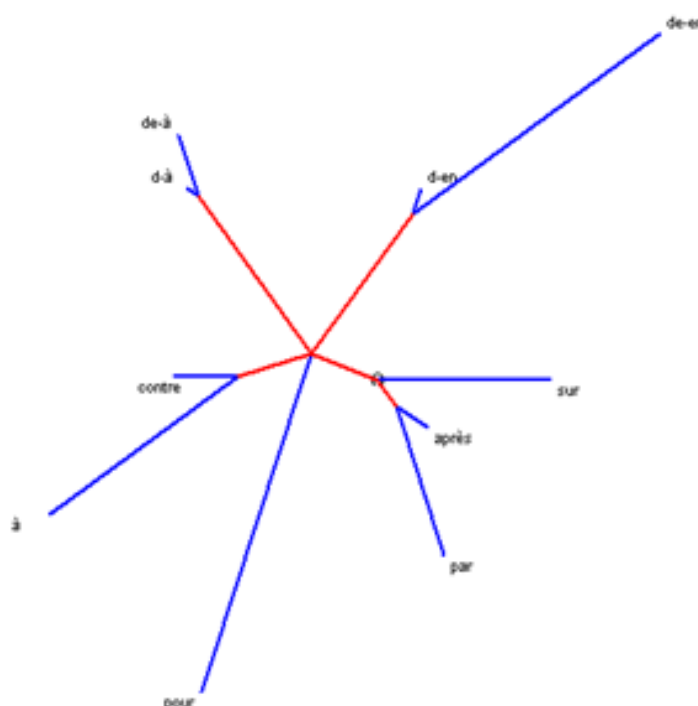
- Les sémantismes qui se réalisent au moyen des structures les plus variées sont l'*Accumulation* et la *Succession* qui sont toutefois sémantiquement proches, liées en particulier par la succession chronologique qu'elles impliquent toutes deux : il y a une quasi-polyvalence des structures, à des degrés divers.

- Chaque sémantisme a une structure privilégiée, sauf la *Succession*, qui s'accommode de manière à peu près équilibrée de la structure avec *à* ou avec *par* ; la primauté d'une structure est moins nette pour l'*Opposition*.

La synonymie des structures est vérifiée par les résultats statistiques mais elle est quantitativement limitée, eu égard aux sémantismes dominants associés à une structure, et ne porte de manière significative que sur deux paires de structures, comme le montre le graphique 1.

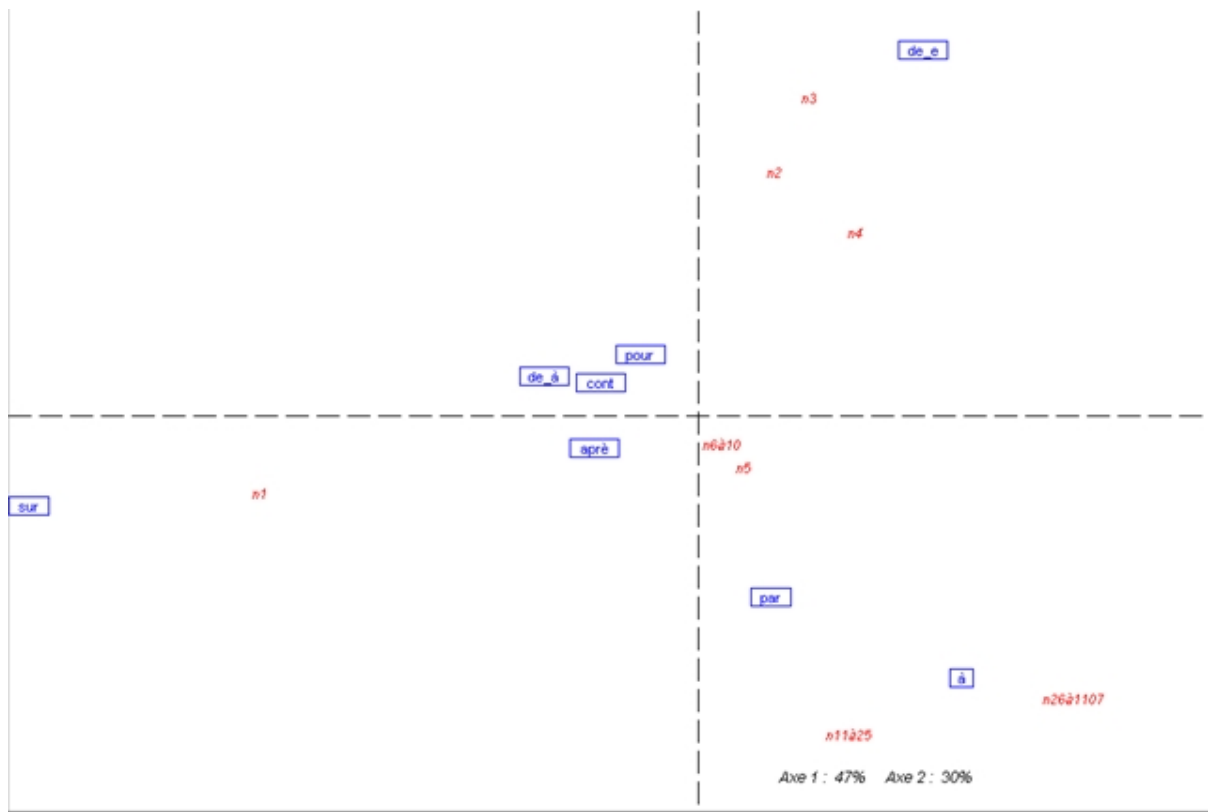
L'analyse arborée montre le rapprochement sur une même branche des prépositions *par* et *après*, *à* et *contre*, signe que les structures N *par* N et N *après* N d'une part et N *à* N et N *contre* N d'autre part se rejoignent par un sémantisme commun.

La polysémie des structures est réelle, comme le montre le graphique 2 : seuls se dégagent les liens *de ... à* – *Relation* et *de ... en* – *Déplacement*. Mais elle est aussi limitée : le graphique arboré 3 montre que partagent surtout les mêmes structures à (*Contact* et *Opposition*) et



pour (Distance et Échange).

Graphique 1 : Les structures selon les sémantismes



Graphique 2 : Analyse factorielle des correspondances qui croise sémantismes et structures

Il y a donc deux phénomènes à observer : d'une part, la polysémie et la polyvalence partielle des structures, d'autre part, la tendance à la spécialisation ; ces deux tendances sont bien marquées sur le graphique 2.

Il est donc possible d'envisager le motif NpN soit à partir des 8 structures, soit à partir des 8 sémantismes (cf. plus loin étude selon les genres).



Graphique 3 : Les sémantismes selon les structures

2. Fréquence et productivité des unités

La fréquence et la productivité des structures sont deux paramètres qui règlent l'appréciation de leur rendement.

- Fréquence

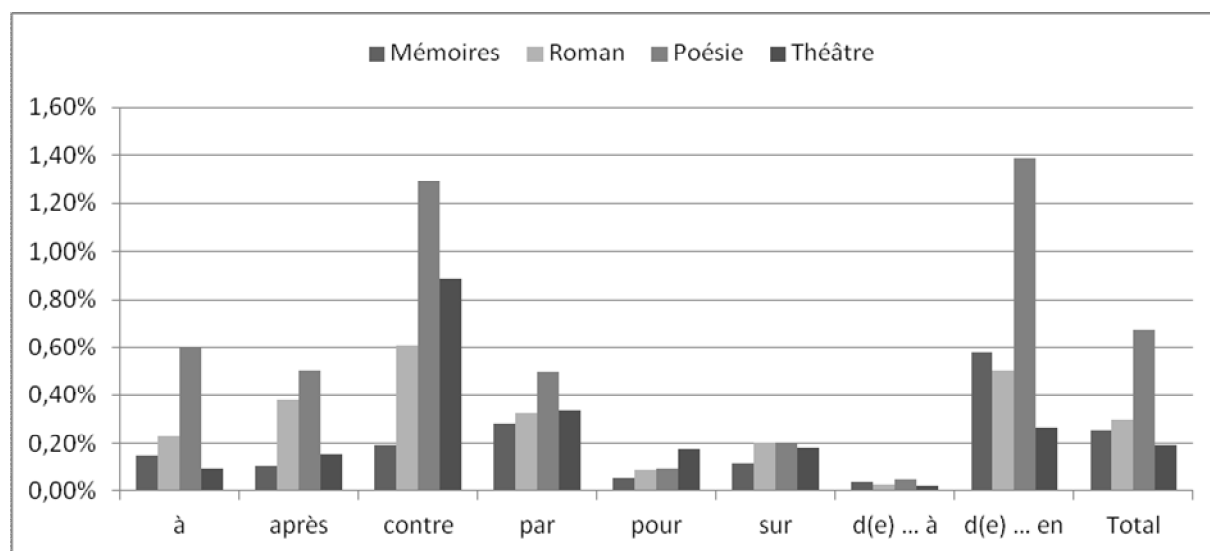
Est calculée la proportion des occurrences des structures NpN sur les effectifs totaux de chaque préposition dans le corpus (*Mémoires Roman Poésie Théâtre* de 1830 à nos jours).

Effectifs des prépositions (A)						Effectifs des structures NPN (B)					
Préposition	Total	Mémoires	Roman	Poésie	Théâtre	Structure	Total	Mémoires	Roman	Poésie	Théâtre
à	1341642	235867	976759	59063	69953	à	2997	354	2219	356	68
après	75393	13171	56448	2590	3184	après	247	14	215	13	5
contre	44305	8392	32101	1781	2031	contre	253	16	196	23	18
par	295617	56121	213080	14505	11911	par	969	160	697	72	40
pour	437191	75488	315216	21746	24741	pour	398	42	292	21	43
sur	349623	52075	255874	26589	15085	sur	663	62	520	54	27
à	1341642	235867	976759	59063	69953	de ... à	372	84	246	27	15
en	772517	124192	570613	39593	38119	de ... en	4222	718	2854	549	101
Total	3316288	565306	2420091	165867	165024	Total	10121	1450	7239	1115	317

Pourcentages des effectifs B dans A					
Structure	Total	Mémoires	Roman	Poésie	Théâtre
À	0,22%	0,15%	0,23%	0,60%	0,10%
après	0,33%	0,11%	0,38%	0,50%	0,16%
contre	0,57%	0,19%	0,61%	1,29%	0,89%
par	0,33%	0,29%	0,33%	0,50%	0,34%
pour	0,09%	0,06%	0,09%	0,10%	0,17%
sur	0,19%	0,12%	0,20%	0,20%	0,18%
de ... à	0,03%	0,04%	0,03%	0,05%	0,02%
de ... en	0,55%	0,58%	0,50%	1,39%	0,26%
Total	0,31%	0,26%	0,30%	0,67%	0,19%

On voit :

- que les prépositions les plus productives en NpN (effectifs) sont *contre* et *en* ;
- que *à* est si fréquent et polyvalent qu’il ne produit pas autant de NpN ;
- que la productivité d’une préposition n’est pas uniforme entre les genres.



Graphique 4 : Les structures selon les genres

Le graphique 4 montre que la poésie est le genre qui mobilise la plus grande partie des effectifs de ses prépositions pour des structures NpN : c’est vrai pour *à*, *après*, *contre*, *par* et *en*.

Ceci est donc un premier indice d’une « fonction » générique de la structure NpN comme possible motif : il est plus spécifique à la poésie.

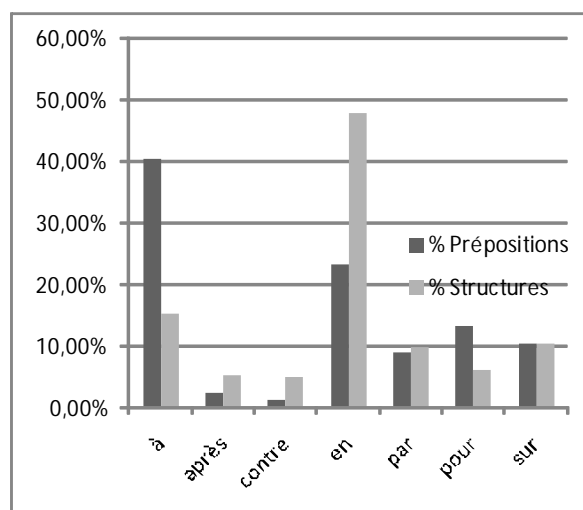
- Productivité

Pour évaluer la productivité des structures, on établit une autre comparaison entre la fréquence des prépositions et le nombre d’expressions différentes produites par les structures NpN. On calcule alors d’une part la proportion des effectifs de chaque préposition sur le total des effectifs des 7 prépositions, d’autre part la proportion des différentes expressions générées par chaque préposition sur le nombre total des expressions

Prépositions			Structures		
à	1341642	40,46%	à / de ... à	347	15,37%
après	75393	2,27%	après	122	5,41%
contre	44305	1,34%	contre	114	5,05%
en	772517	23,29%	de ... en	1083	47,98%
par	295617	8,91%	par	220	9,75%
pour	437191	13,18%	pour	137	6,07%
sur	349623	10,54%	sur	234	10,37%
Total	3316288			2257	

Graphique 5 : La productivité des structures

À et *en* sont les deux prépositions les plus fréquentes dans le corpus, mais leur produc-



tivité est inverse : à génère nettement moins

d'expressions différentes que *en* dans la structure *de ... en*. Comme pour *à*, les autres écarts sont le signe d'une spécialisation partielle des structures qui confine à la lexicalisation ou au figement : *œil pour œil, corps à corps*, ou, à l'inverse, d'une productivité : *après, contre*.

4.3. La création lexicale

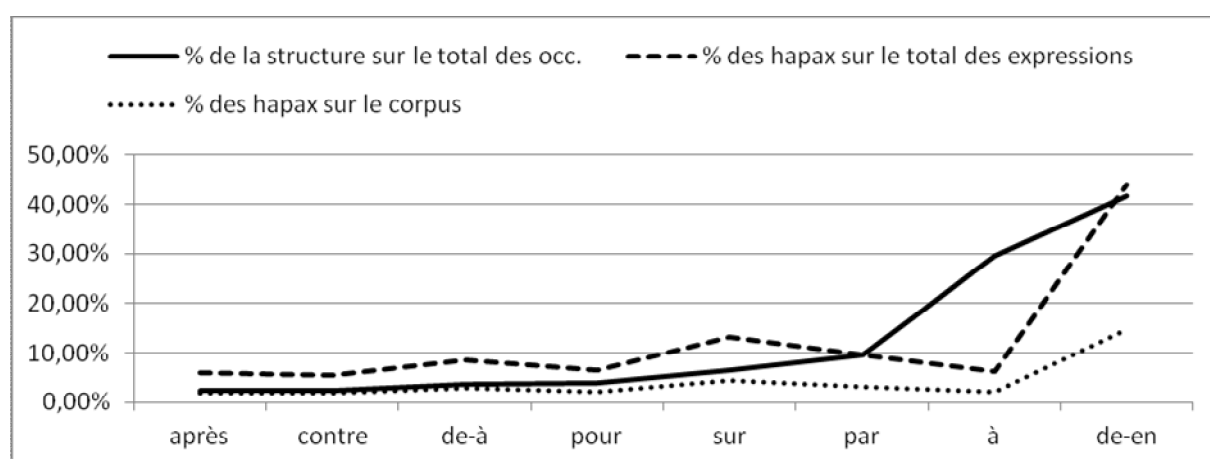
La création de nouvelles expressions est un fait de discours susceptible d'être interprété comme fait de style. Elle peut se faire par analogie à partir de syntagmes figés en langue. Par exemple, si *côte à côte* est identifié comme locution adverbiale par le TLF, *frange à frange* est attesté en discours mais n'apparaît pas dans l'article *frange* du dictionnaire. Le sens de l'expression se comprend par référence à d'autres expressions structurellement équivalentes et usuelles et par sa contextualisation, mais l'expression n'est pas grammaticalisée. Un indice de la productivité des structures est la fréquence des hapax dans un corpus donné.

Faute de pouvoir établir facilement et avec certitude quelles structures existent en langue et lesquelles sont des créations, partons d'un critère statistique sévère :

- les plus fréquentes sont forcément figées jusqu'à se grammaticaliser pour certaines et à fonctionner le plus souvent comme locutions adverbiales ;
- moins une expression est attestée, plus elle échappe à la lexicalisation.

De ce postulat, on déduira que les hapax sont des indices de productivité ; pour une structure, compter le nombre d'expressions de fréquence 1 dans le corpus établit une nouvelle proportion entre le nombre d'hapax et le total des expressions de la structure. Le comptage se fait sur le « vocabulaire » : combien d'expressions différentes sont hapax et combien ne sont pas hapax ?

On pourrait penser que la proportion des expressions hapax est directement liée à la fréquence des structures : plus une structure est fréquente, plus le nombre de ses occurrences qui sont des hapax est grand. C'est effectivement le cas, mais avec les exceptions qui indiquent une plus grande productivité (autrement dit une plus grande capacité à produire des hapax) pour une structure et le phénomène inverse pour une autre :



Graphique 6 : La productivité des structures et les hapax

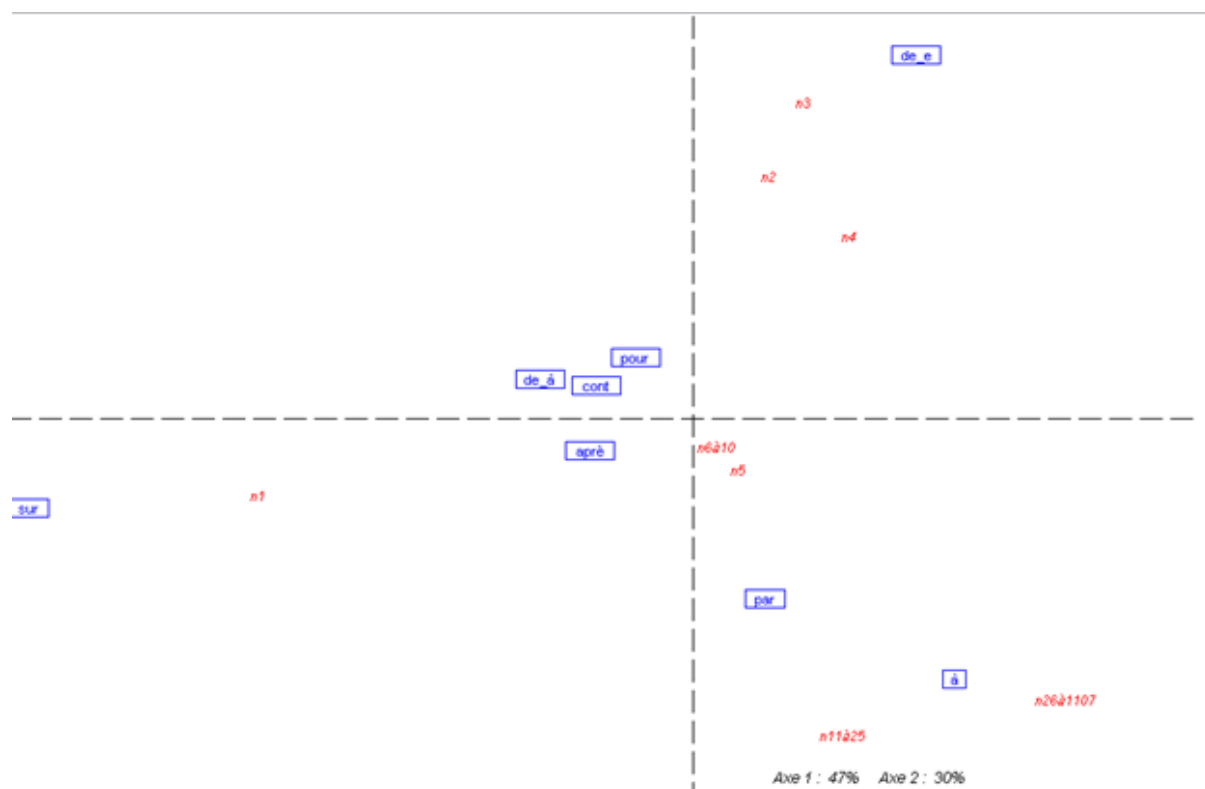
Le graphique 6 est calculé à partir de :

- la proportion de chaque structure sur le total des occurrences des NpN, ce qui donne un ordre de grandeur, de la moins fréquente (*après*) à la plus fréquente (*de ... en*) ;
- la proportion des hapax d'une structure sur le total de ses expressions, sans tenir compte de leurs effectifs : chaque expression est comptée pour une, et le calcul porte sur les hapax parmi celles-ci ;
- la proportion des hapax de chaque expression dans tout le corpus des NpN.

On voit que les courbes sont assez parallèles, mais :

- la structure *sur* produit plus d'hapax que ne le laisse attendre sa fréquence absolue, comparée aux autres structures ;
- la structure *à* accuse un déficit en hapax : elle est moins productive et plus riche en expressions figées ;

Le graphique AFC ci-dessous confirme ces observations :



Graphique 7 : AFC des hapax et non-hapax par structures

Les hapax (n1) sont du même côté que *sur*, la structure *à* est dans le même quadrant que les expressions à haute fréquences.

5. Les genres

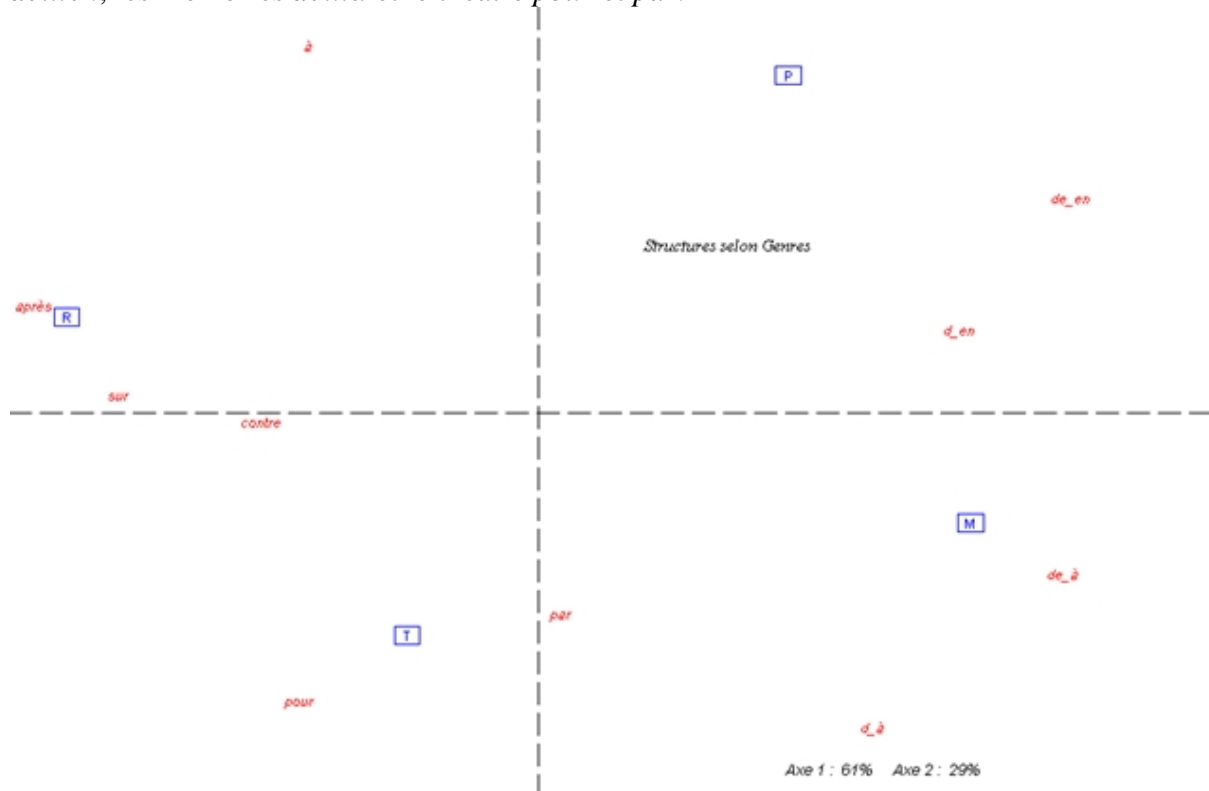
Les paramètres observés jusqu'à présent, sémantisme et productivité, méritent d'être affinés par une approche de la population des structures NpN selon les genres.

Les effectifs globaux de chaque structure sont donnés dans les premiers tableaux.

5.1. La productivité des structures selon les genres

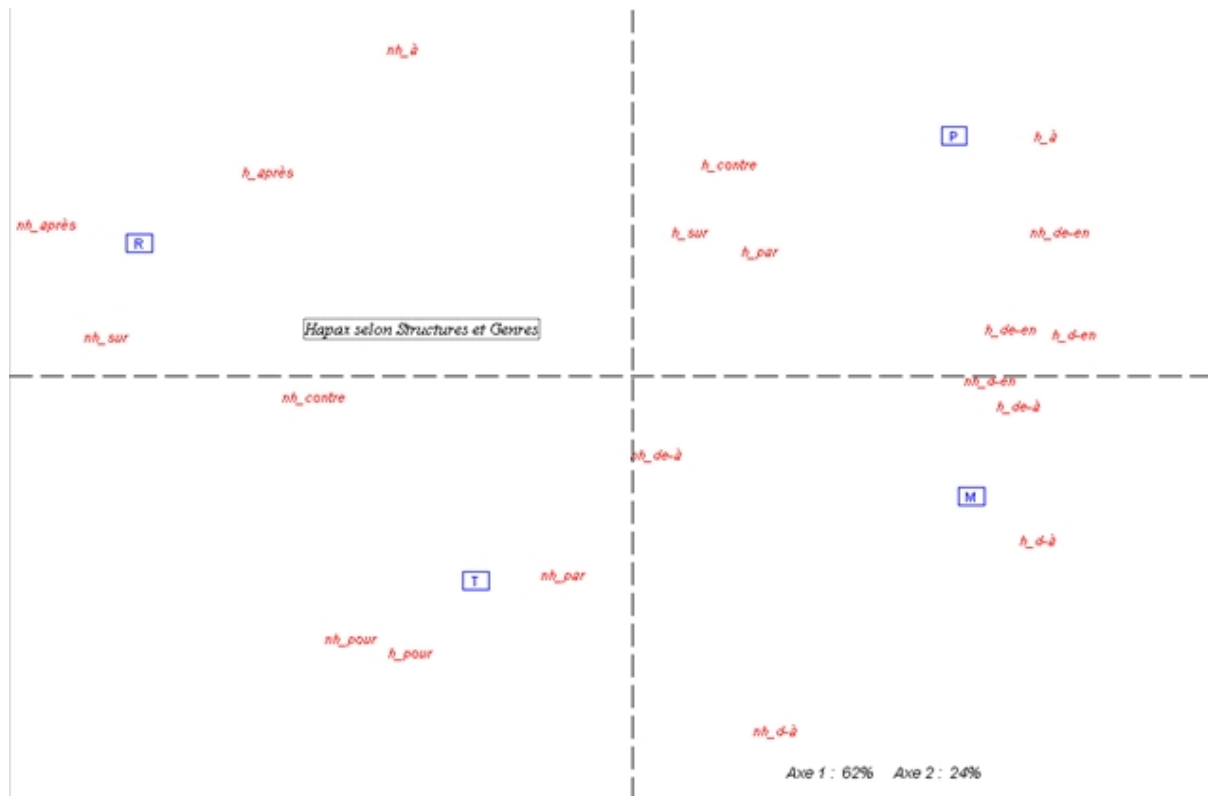
Certains genres entretiennent des affinités avec certaines structures (graphique 8) :

Autour du roman gravitent les structures *après*, *sur* et *contre*, tandis que la poésie attire *de...en*, les mémoires *de...à* et le théâtre *pour* et *par*.



Graphique 8 : Les structures selon les genres

L'observation vaut surtout si l'on détaille les données en distinguant hapax et non-hapax (graphique 9) :



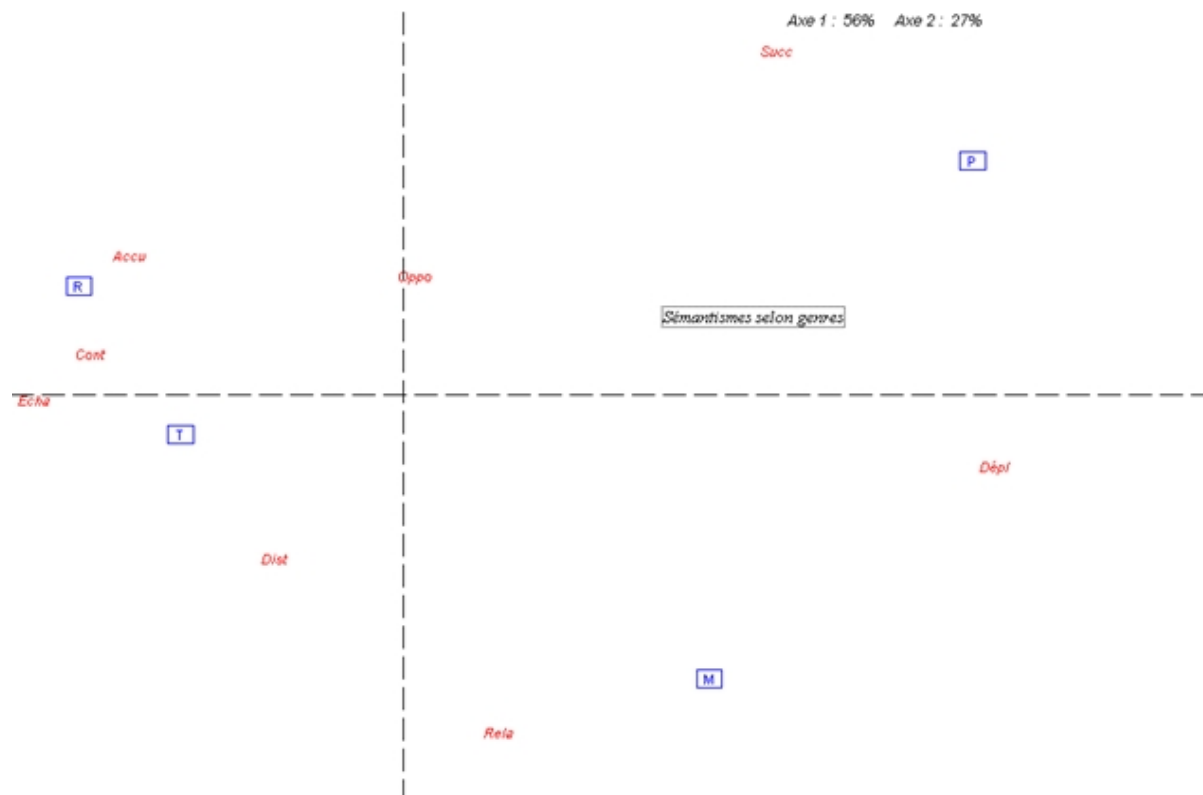
Graphique 9 : Hapax et non-hapax selon les genres

On voit que la plupart des structures, dans les occurrences d'expressions hapax et non-hapax, restent en place par rapport aux genres, à l'exception de la poésie, qui attire les occurrences hapax de plusieurs structures : *contre*, *sur*, *par*, *à*, et *de ... en*.

Ceci constitue un deuxième indice de l'affinité d'un genre parmi les autres pour le motif NpN : la poésie est le genre le plus propice à la création d'expression nouvelles à partir des structures *contre*, *sur*, *par*, *à* et *de...en*.

5.2. Les sémantismes et les genres

L'affinité des genres et des sémantismes est représentée dans le graphique 10.



Graphique 10 : Les sémantismes selon les genres

Cette AFC ne double qu'en partie celle des structures, en fonction des sémantismes privilégiés de chacune d'entre elle. On peut supposer que le roman est caractérisé par le sémantisme du *Contact* en raison de son affinité avec la structure *contre*. La poésie se rapproche de la *Succession* – sémantisme réalisé par le plus grand nombre de structures sans qu'aucune n'affirme sa prédominance. C'est bien un sémantisme propre à la poésie.

Conclusion

À la question qui était posée en introduction, à savoir la recevabilité de la candidature de la structure NpN au rang de motif, la réponse est positive. Les critères de fréquence en langue et de récurrence dans le corpus sont vérifiés. La mémorisation est aisée, soit que l'expression soit lexicalisée, soit que le lecteur soit apte à reconnaître un pattern syntaxique par analogie avec les expressions figées. La productivité à partir de schémas préconstruits est variable mais puissante, tout en évoluant autour de huit sémantismes identifiés. De fait, ce motif se situe entre langue et discours et se développe comme motif discursif susceptible de produire un effet stylistique corrélé à l'innovation lexicale, dont les hapax sont un indice. Le motif peut jouer un rôle dans la caractérisation générique, puisqu'a été observée une répartition claire, même si elle ne peut être complètement interprétée ici, entre les genres. La poésie, notamment, se caractérise par les observations suivantes : c'est le genre qui mobilise la plus grande part des effectifs de ses prépositions dans des structures NpN ; c'est le genre le plus propice à la création d'expressions nouvelles à partir des structures *contre*, *sur*, *par*, *à*, *de...en* et se trouve associée au sémantisme de la succession. Ces remarques strictement factuelles sont des amorces à d'autres études plus linguistiques qui pourront être développées afin d'affiner et d'interpréter ces résultats.